CLAIMS

What is claimed is:

1. A method for predicting the secondary structure of proteins comprising (a) obtaining a multiplicity of homologous protein sequences, (b) constructing an alignment of the multiplicity of sequences, and (c) analyzing patterns of conservation and variation at sites in the multiple sequence alignment, wherein said multiplicity comprises at least 16 homologous protein sequences.

2. The method of Claim 1, wherein said set comprises at least eight pairs of proteins, wherein the proteins in each pair are at least 80% identical in sequence.

3. The method of Claim 1, wherein said analysis incorporates a model for the evolutionary divergence of said homologous protein sequences.

4. A method for the identification of a secondary structural element that may be involved in functional adaptation, wherein said method comprises (a) obtaining a multiplicity of homologous protein sequences and their encoding DNA sequences, (b) constructing an alignment of the multiplicity of sequences, (c) constructing an evolutionary tree that models the evolutionary history of the family of genes and proteins represented by said sequences, (d) constructing models of the sequences of the genes and their encoded proteins at nodes in the tree, (e) assigning changes in the gene and protein sequences to lines connecting such nodes, and (f) calculating the ratio of non-synonymous to synonymous nucleotide substitutions for said lines at sites in said alignment that are part of said element, wherein said secondary structural element is identified as possibly being involved in functional adaptation if the said ratio is in excess of a preselected value.

5. A method for identifying a pair of proteins that may come into physical contact when they function comprising (a) obtaining a multiplicity of homologous protein sequences and their encoding DNA sequences that are related to each member of the pair, (b) constructing an alignment of the multiplicity of sequences, (c) constructing an evolutionary tree that models the evolutionary history of the family of genes and proteins represented by said sequences, (d) constructing models of the sequences of the genes and their encoded proteins at nodes in the tree, and (e) assigning events in the gene and protein sequences to lines connecting such nodes, wherein said pair of proteins is identified as possibly coming into physical contact when they function if events assigned to a line in one family correlate with events assigned to lines representing contemporaneous episodes in the other family.

6. The method of Claim 5, wherein said events comprise episodes of sequence evolution associated with a ratio of non-synonymous to synonymous nucleotide substitutions in excess of a preselected value.

7. The method of Claim 5 wherein one protein in said pair is a peptide hormone, and the other protein in said pair is a peptide hormone receptor.

8. A method for estimating the date since a pair of proteins diverged comprising aligning the sequences of said pair, identifying in said alignment each cysteine, aspartic acid, glutamic acid, phenylalanine, histidine, lysine, asparagine, glutamine, and tyrosine that is conserved in the pair, totalling the number of these, and summing the number of these wherein the respective codon is conserved, obtaining a ratio by dividing said sum by said total, subtracting 0.5 from said ratio, multiplying the difference by 2, taking the natural logarithm of the product and dividing by a number that is the estimate for the first order rate constant for replacement at the silent sites in said codons.

9. A method for identifying a protein family that may be associated with a change in a physiology in a taxon, said method comprising (a) obtaining a multiplicity of homologous protein sequences for said family, (b) constructing an alignment of the multiplicity of sequences, (c) constructing an evolutionary tree that models the evolutionary history of said family, and (d) correlating events in said family in time with the change in said physiology.

10. The method of Claim 9, wherein said time is estimated using the paleontological record.

11. The method of Claim 9, when dating events in the evolutionary history of said family is done using the method of Claim 8.

12. The method of Claim 9, wherein said events comprise gene duplications.